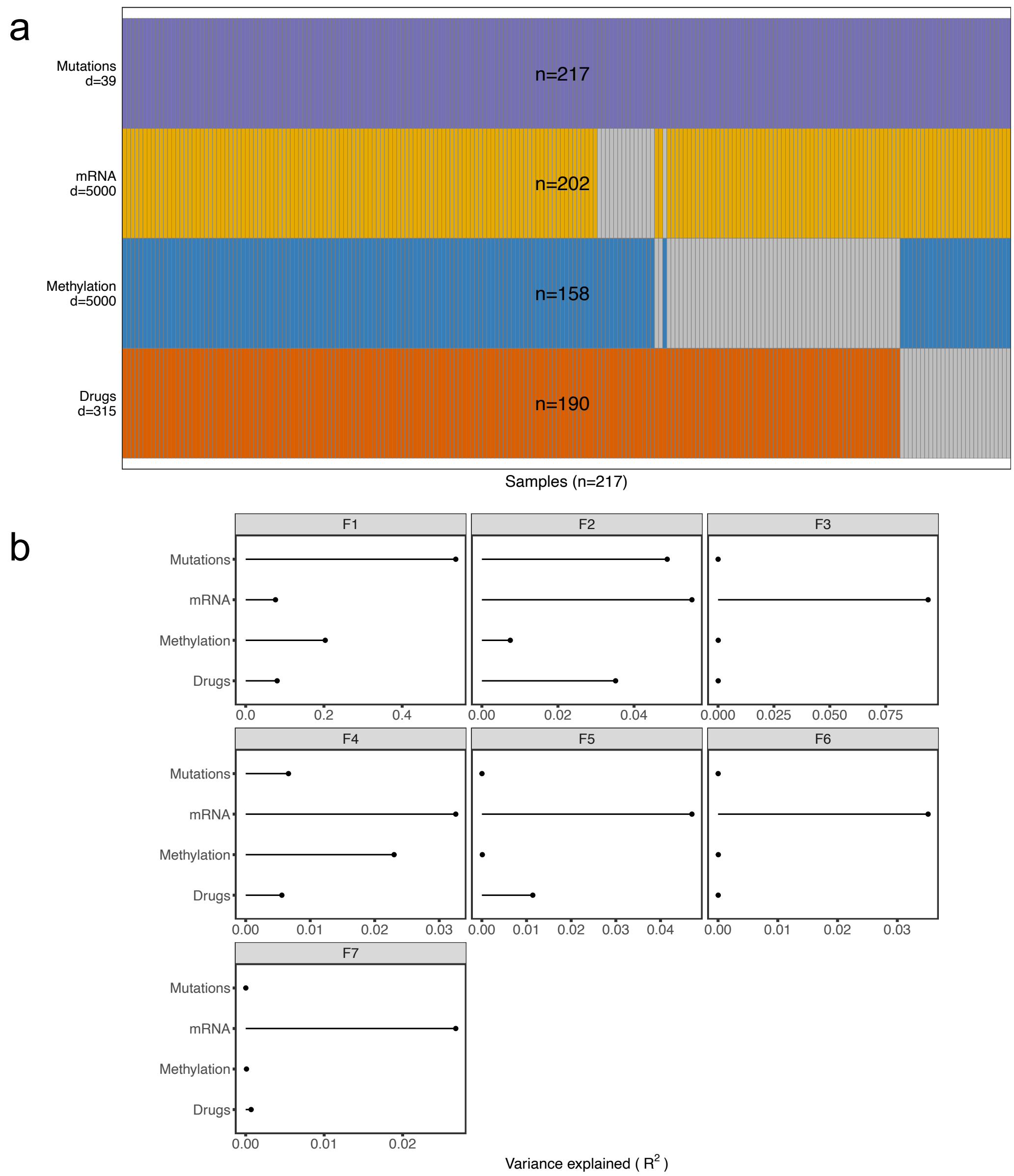
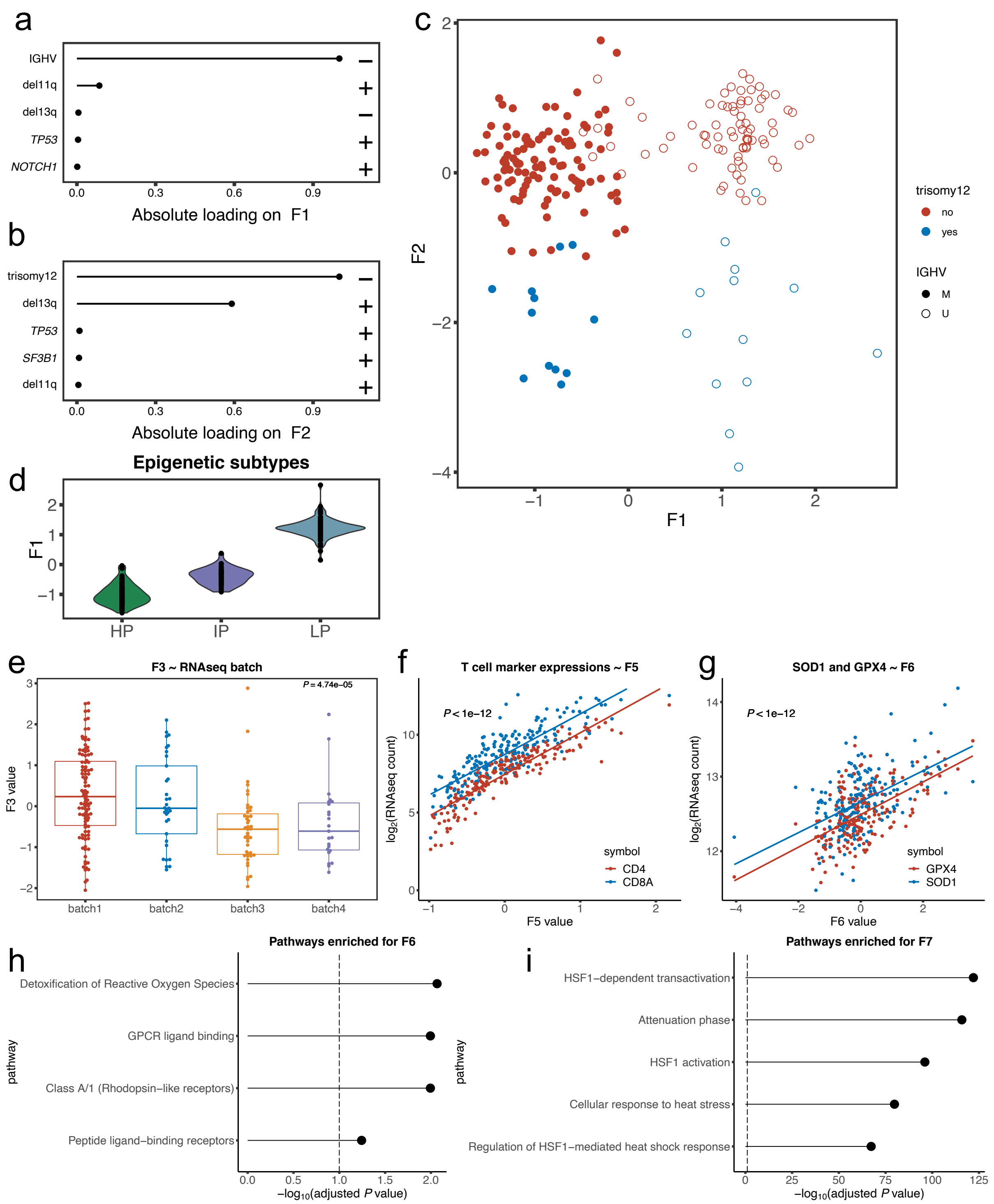
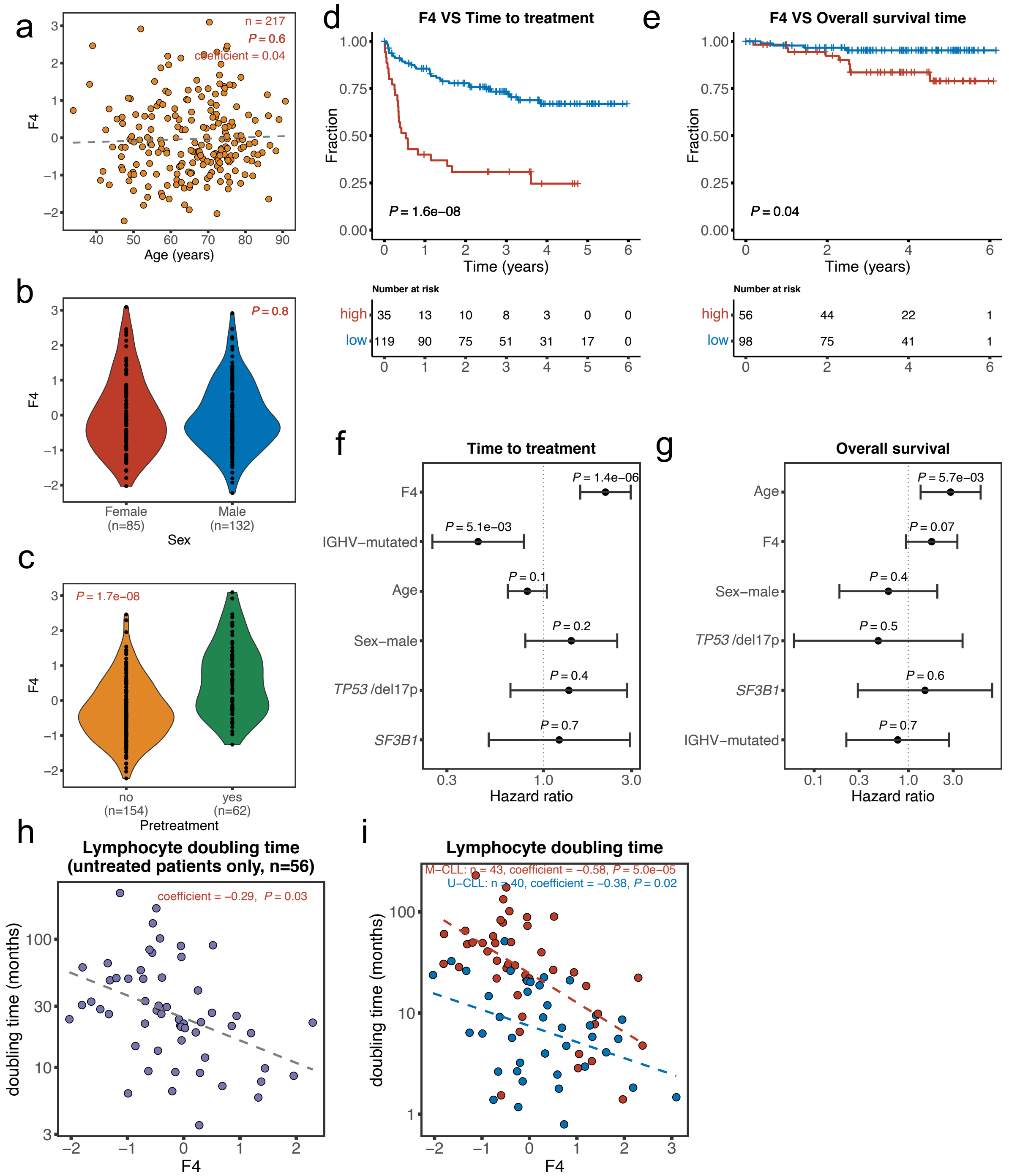


Extended data figure 1. Integration of multi-omics profiling datasets using multi-omics factor analysis (MOFA). a, Datasets included in the MOFA training model and the overlap of patient samples among datasets. The number of features in each dataset is indicated by “d=” and the number of samples in each dataset is indicated by “n=”. **b,** Stem plots showing the variance explained (R^2) values for each view by each factor.

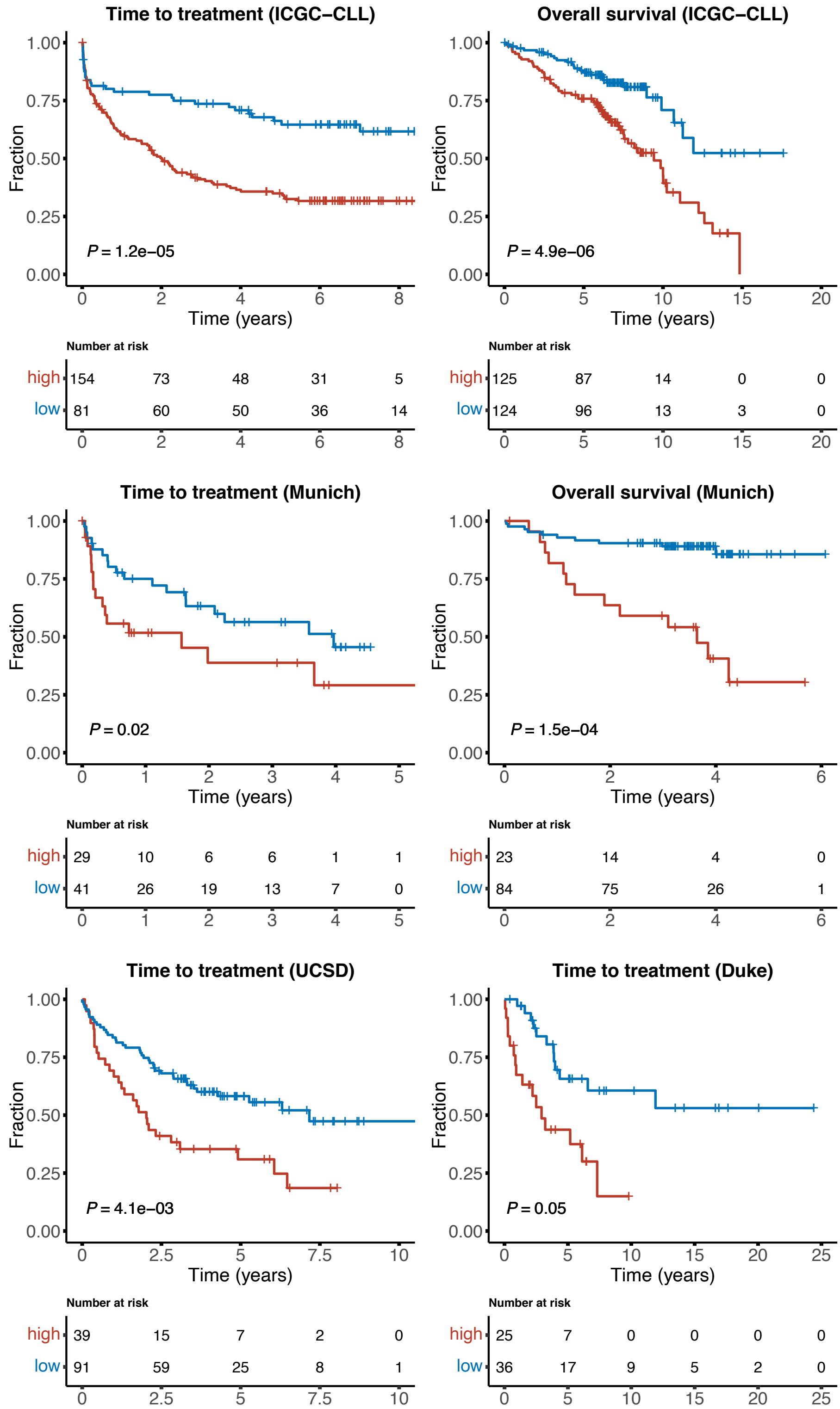


Extended data figure 2. Characterization of the factors identified by MOFA.
a and b, Absolute loadings of the top features of F1 and F2 in the genomic dataset (*n*=217 samples). **c**, Visualization of patient samples using F1 and F2 as coordinates. A dot represents a primary CLL with mutated IGHV status (M-CLL, *n*=117 samples), and a circle represents a primary CLL with unmutated IGHV status (U-CLL, *n*=89 samples). CLL with (*n*=25 samples) and without trisomy12 (*n*=181 samples) are colored by blue and red, respectively. **d,** Association between F1 and three epigenetic subtypes of CLL: HP (high-programmed, *n*=86 samples), IP (intermediate-programmed, *n*=35 samples) and LP (low-programmed, *n*=86 samples). F1 separated the three epigenetic subtypes in their proper order (HP-, IP- and LP-CLL). **e,** F3 values for CLL samples in different RNAseq batch (*n*=103, 33, 43 and 23 samples

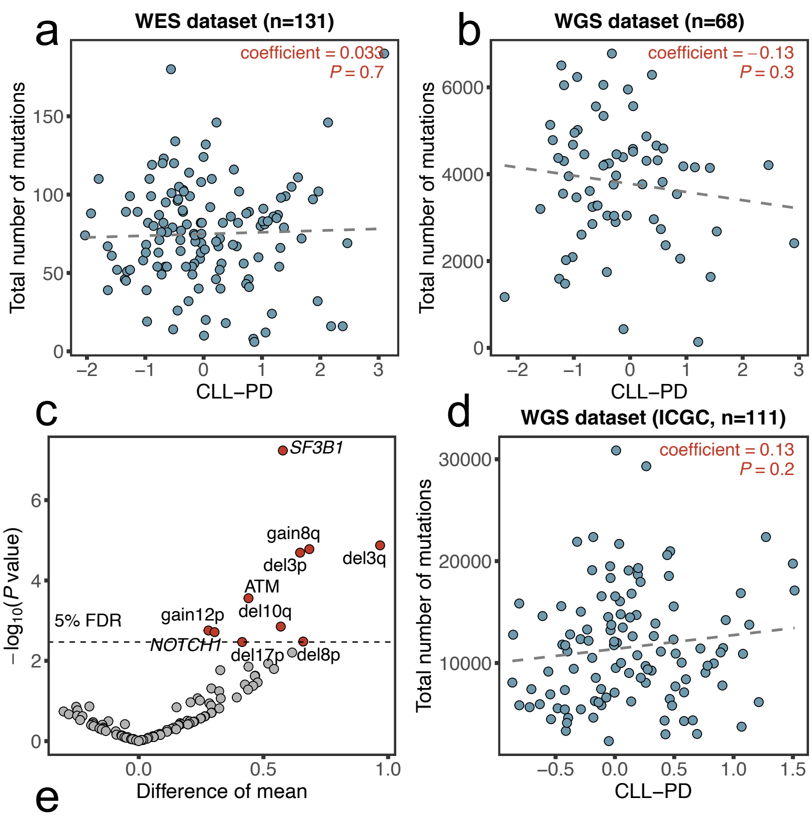
for batch 1, 2, 3 and 4, respectively). Each dot represents a patient sample. The boxplot shows the interquartile range in the box with the median as a horizontal line. Whiskers extend to 1.5 times the interquartile range. *P* value was calculated by ANOVA test. **f,** Correlations between Factor 5 and the mRNA expression of T cell markers genes: CD4 and CD8A. *P* values are from two-sided Pearson's correlation tests. **g,** Correlations between Factor 6 and the expression of two exemplary genes (SOD1 and GPX4) involved in the response to reactive oxygen species (ROS). *P* values are from two-sided Pearson's correlation tests. **h,** Pathway enrichment results for Factor 6. Enrichment *P* values were adjusted by Benjamini-Hochberg method. **i,** Pathway enrichment results for Factor 7. Enrichment *P* values were adjusted by Benjamini-Hochberg method. Factor 5 and Factor 7 were characterized in detail, under the names of Factor 4 and Factor 5 respectively, in the article describing the implementation of MOFA[18]. All analysis results shown in panel **f - i** were performed on RNAseq data from 202 samples.



Extended data figure 3. **Associations between Factor 4 and demographic and clinical characteristics. a,** Association of F4 to age. *P* values is from two-sided Pearson's correlation test. (*n*=217 samples) **b and c,** Associations of F4 to sex and pretreatment status. *P* values are from two-sided t-tests. **d and e,** Kaplan-Meier plots for showing the associations between F4 and TTT or OS in patients without previous treatment. The *P*-values were assessed by Cox regression models with F4 as a continuous variable. For visualization purposes only, optimal cutoffs to separate patients into high and low CLL-PD groups were estimated by the maximally selected rank test implemented in the R/CRAN package maxstat (v0.7). **f and g,** Forest plots showing the hazard ratios with 95% confidence intervals and *P* values from multivariate Cox models that include known demographic and genomic risk factors, for TTT and OS in patients without previous treatment. F4 remained significantly associated with TTT in multivariate analysis. In multivariate analysis for OS, none of the risk factors except for age were significant, however, the hazard ratio showed the same trend for F4 as in the full data set analysis, consistent with the reduced statistical power of the subset analysis. (*n*=154 patients) **h,** Correlation between F4 and lymphocyte doubling time (LDT) in previously untreated patients. *P* values and coefficients are from two-sided Pearson's correlation tests. **i,** Correlation between F4 and lymphocyte doubling time (LDT) in M/U-CLL separately. *P* values and coefficients were from two-sided Pearson's correlation tests. (*n*=43 and 40 samples for M-CLL and U-CLL, respectively).



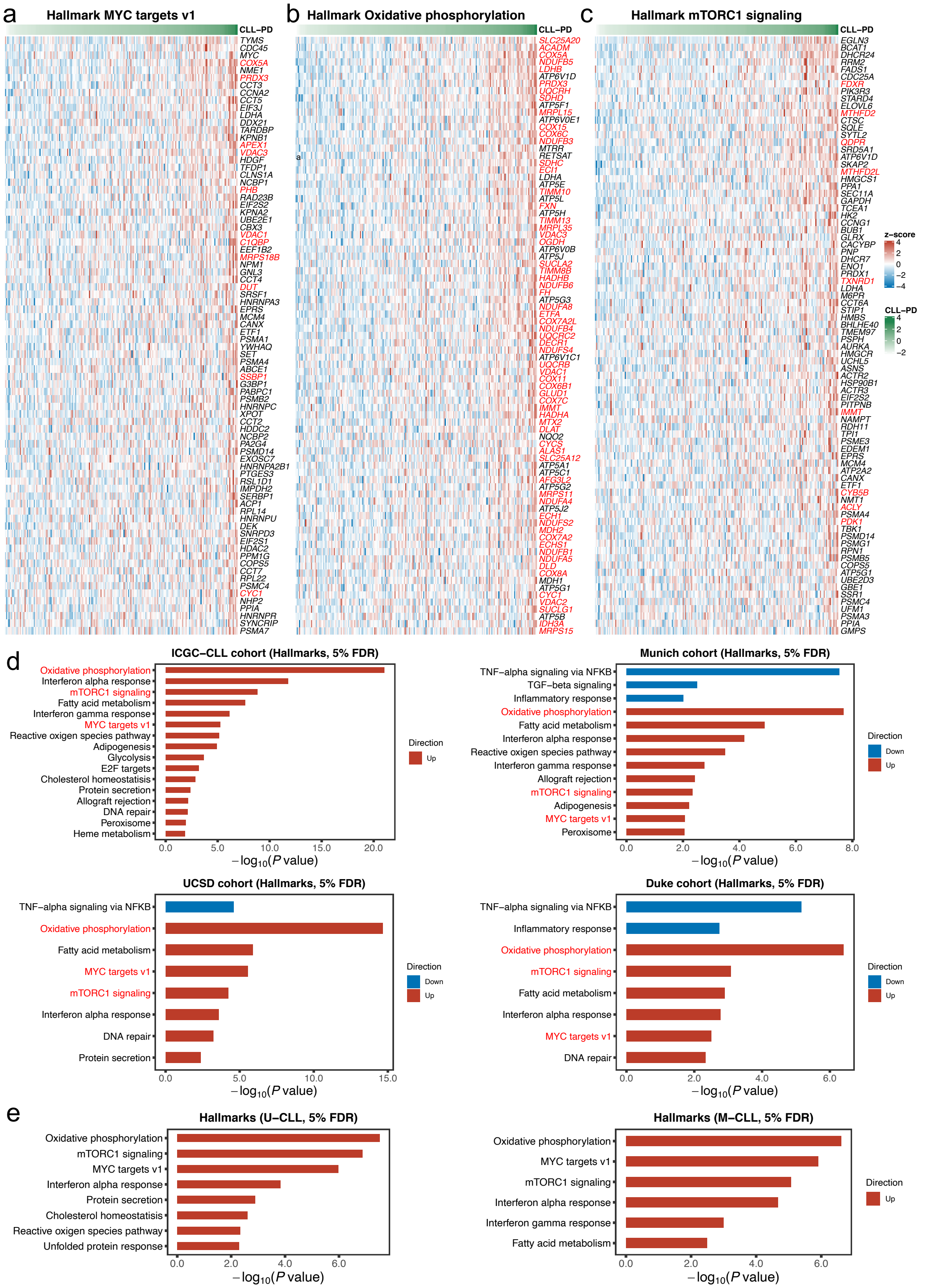
Extended data figure 4. Associations between CLL-PD score and outcomes (TTT or OS) in four external CLL cohorts with gene expression data. The per-test *P*-values were calculated by two-sided log-rank tests on Cox regression models with CLL-PD score as a continuous variable. For visualization purposes only, optimal cutoffs to separate patients into high and low CLL-PD groups were estimated by the maximally selected rank test implemented in the R/CRAN package *maxstat (v0.7)*.



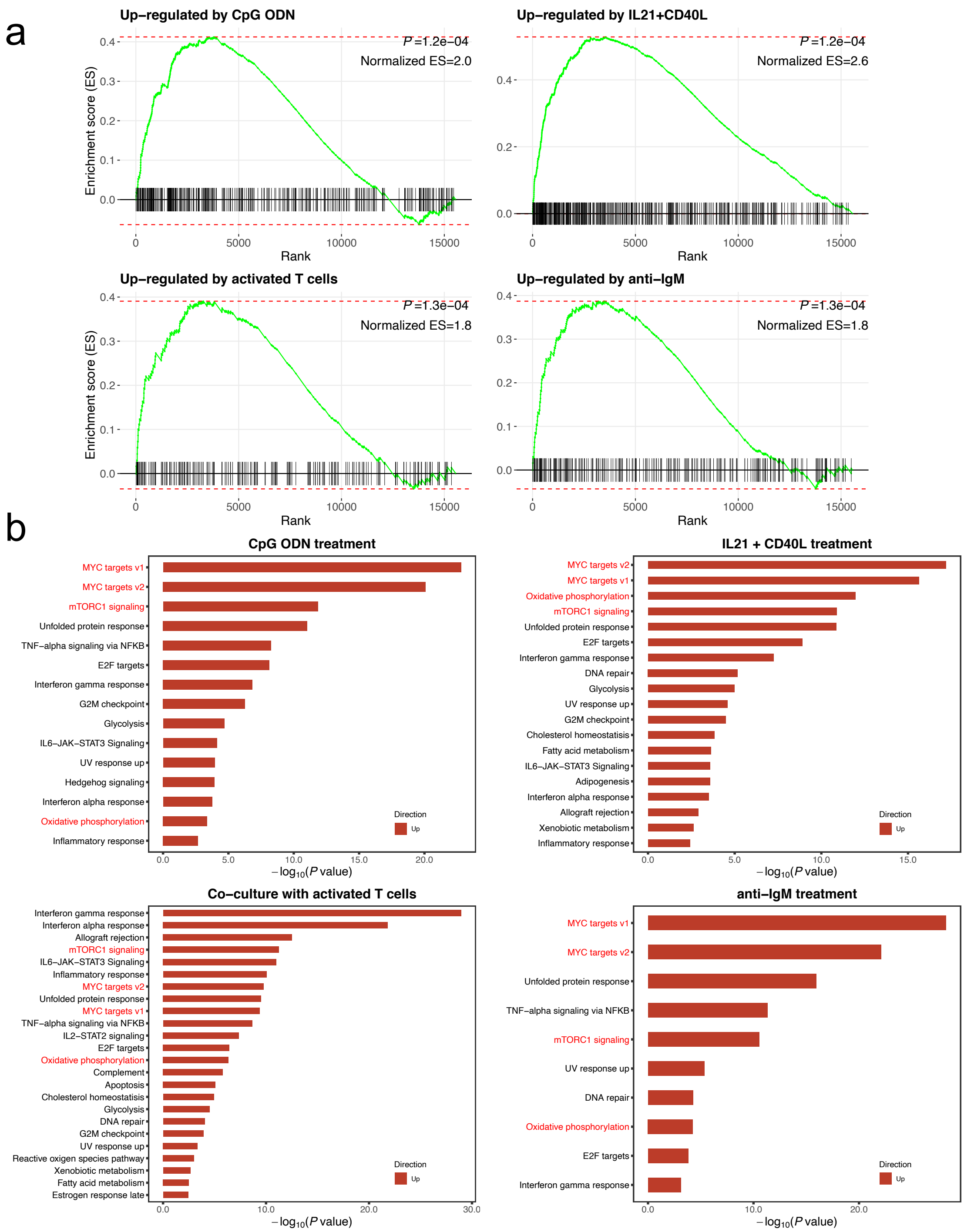
e

Rank	Motif	P-value	log P-pvalue	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1	GGCCACATGGCC	1e-19	-4.489e+01	32.64%	18.74%	134.2bp (197.9bp)	MYCN/MA0104.4/Jaspar(0.760) More Information | Similar Motifs Found	motif file (matrix)
2	CTTCATGGGGCT	1e-19	-4.453e+01	14.24%	5.29%	143.1bp (206.3bp)	CTCFL/MA1102.1/Jaspar(0.720) More Information | Similar Motifs Found	motif file (matrix)
3	GCAGGACAGCCA	1e-18	-4.355e+01	19.68%	8.99%	136.6bp (196.7bp)	ZNF341(Zf)/EBV-ZNF341-ChIP-Seq(GSE113194)/Homer(0.615) More Information | Similar Motifs Found	motif file (matrix)
4	CACATCAGGA	1e-17	-4.045e+01	10.88%	3.58%	119.4bp (204.6bp)	PB0207.1_Zic3_2/Jaspar(0.736) More Information | Similar Motifs Found	motif file (matrix)
5	CTGACACCAACA	1e-17	-3.963e+01	6.48%	1.36%	133.1bp (227.8bp)	PH0164.1_Six4/Jaspar(0.756) More Information | Similar Motifs Found	motif file (matrix)
6	GTGCCCACAG	1e-16	-3.816e+01	43.29%	29.03%	140.6bp (212.4bp)	HIC2/MA0738.1/Jaspar(0.739) More Information | Similar Motifs Found	motif file (matrix)
7	CATGGAGCAC	1e-15	-3.680e+01	35.65%	22.54%	145.2bp (201.7bp)	Znf263(Zf)/K562-Znf263-ChIP-Seq(GSE31477)/Homer(0.726) More Information | Similar Motifs Found	motif file (matrix)
8	GTGGAGCAGCTG	1e-15	-3.529e+01	4.17%	0.58%	128.0bp (119.6bp)	ASCL1/MA1100.1/Jaspar(0.878) More Information | Similar Motifs Found	motif file (matrix)
9	GGTGCCATTC	1e-15	-3.528e+01	13.08%	5.32%	145.1bp (218.0bp)	PH0164.1_Six4/Jaspar(0.729) More Information | Similar Motifs Found	motif file (matrix)
10	CACACCTGGG	1e-15	-3.498e+01	26.04%	14.84%	142.5bp (216.8bp)	ZEB2(Zf)/SNU398-ZEB2-ChIP-Seq(GSE103048)/Homer(0.789) More Information | Similar Motifs Found	motif file (matrix)

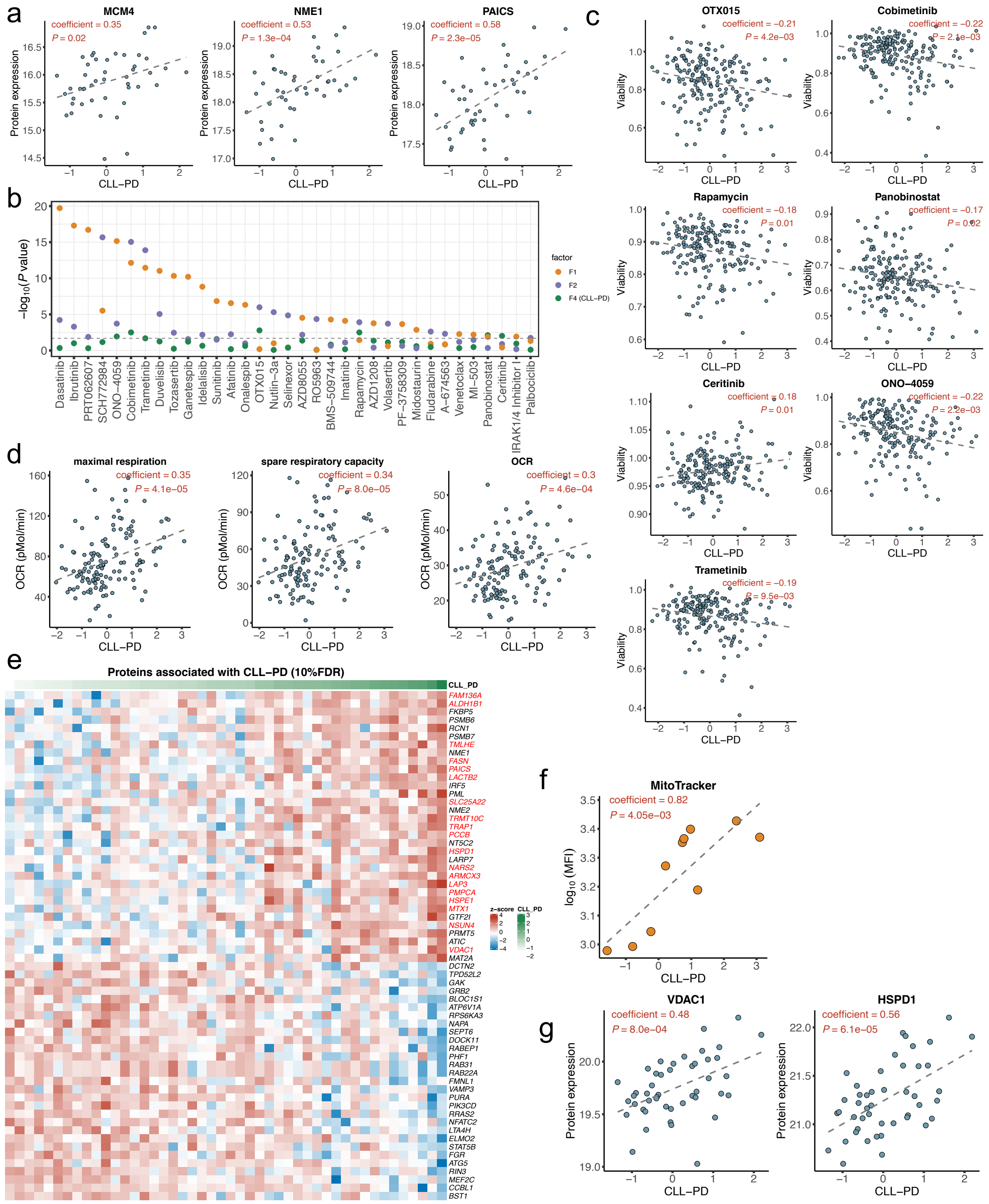
Extended data figure 5. Associations of CLL-PD to genomic aberrations and DNA methylation. a and b, Scatter plots showing the associations between CLL-PD and the total number of mutations detected by whole exome sequencing **(a)** or whole genome sequencing **(b).** Mutations on immunoglobulin genes were excluded when calculating the total number of mutations to avoid potential influence of somatic hypermutation. *P* values and coefficients were calculated by two-sided Pearson's correlations tests. **c,** Associations of the CLL-PD score to genomic aberrations in the ICGC-CLL cohort (*n*=249 samples). *P* values are from two-sided t-tests. **d,** Associations of the CLL-PD score to overall mutation load in the ICGC-CLL cohort. *P* value is from two-sided Pearson's correlation test. **e**, top 10 enriched transcription factor binding motifs in the regions that show hypomethylation in samples with high CLL-PD values, *P* values were calculated by the Homer de novo algorithm[32].



Extended data figure 6. Gene expression signatures of CLL-PD. a to c, Heatmap plots showing the z-score of the expression values of genes that are significantly correlated with CLL-PD (1% FDR, Benjamini-Hochberg's method) and are in the Hallmark MYC targets v1 (a), Hallmark oxidative phosphorylation (OXPHOS) (b) and Hallmark mTORC1 signaling (c) gene sets from Molecular Signatures Database (MSigDB)[39]. Samples (columns) are ordered by their CLL-PD values. Symbols of the genes coding mitochondrial proteins are colored in red. **d**, Gene enrichment analysis of genes correlated with the CLL-PD scores in the four external cohorts shown in Figure 2b, using Hallmark gene sets from MSigDB. The names of gene sets related to MYC targets, mTOR signaling and OXPHOS are colored in red. (*n*=249, 107, 130 and 81 patients for the ICGC-CLL, Munich, UCSD and Duke cohorts, respectively) **e,** Gene set enrichment analysis of genes correlated with CLL-PD in U-CLL (*n*=107 samples) and M-CLL (*n*=93 samples) separately.

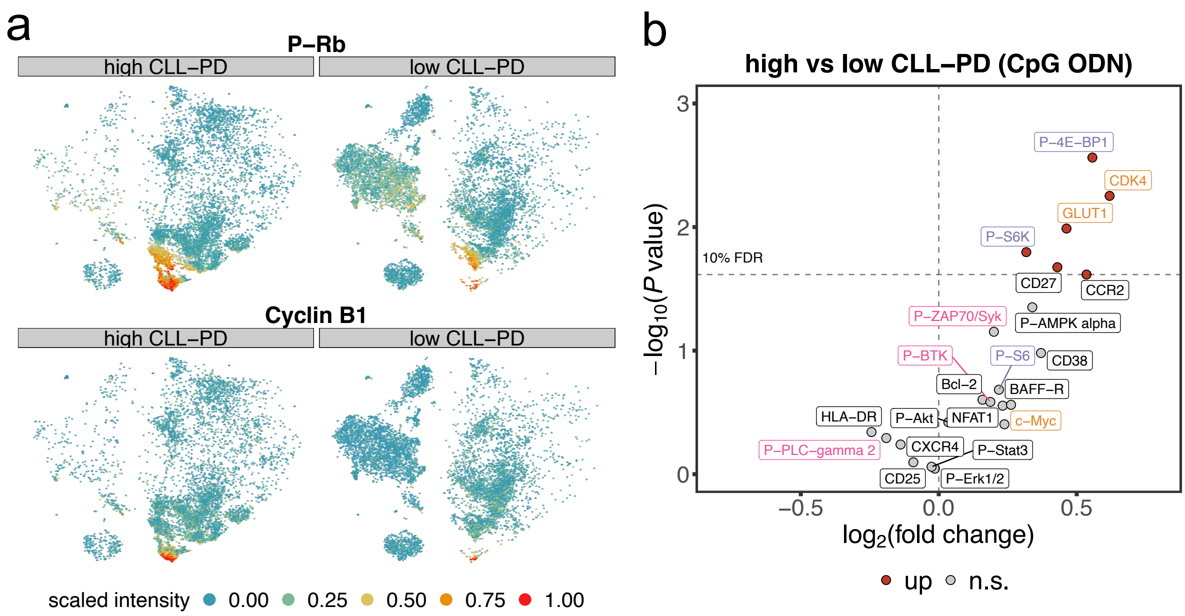


Extended data figure 7. Comparison between the gene expression signatures of CLL-PD and the signatures of pro-proliferative stimuli. a, GSEA plots showing the enrichment of CLL-PD correlated genes in the gene sets defined on the genes significantly up-regulated by the four indicated pro-proliferative stimuli (1% FDR and log2 fold change >1). **b,** Gene enrichment analysis of genes differentially regulated after four pro-proliferative microenvironment stimulations: including CpG ODN (ArrayExpress ID: E-GEOD-30105, *n*=9 samples), co-culturing with T-cells (ArrayExpress ID: E-GEOD-50572, *n*=5 samples), IL21+CD40L (ArrayExpress ID: E-GEOD-50572, *n*=4 samples), and cross-linked anti-IgM (ArrayExpress ID: E-GEOD-39411, *n*=11 samples). Gene sets that passed a threshold corresponding to an FDR of 5% are shown. The names of gene sets related to MYC targets, mTOR signaling and OXPHOS are colored in red.

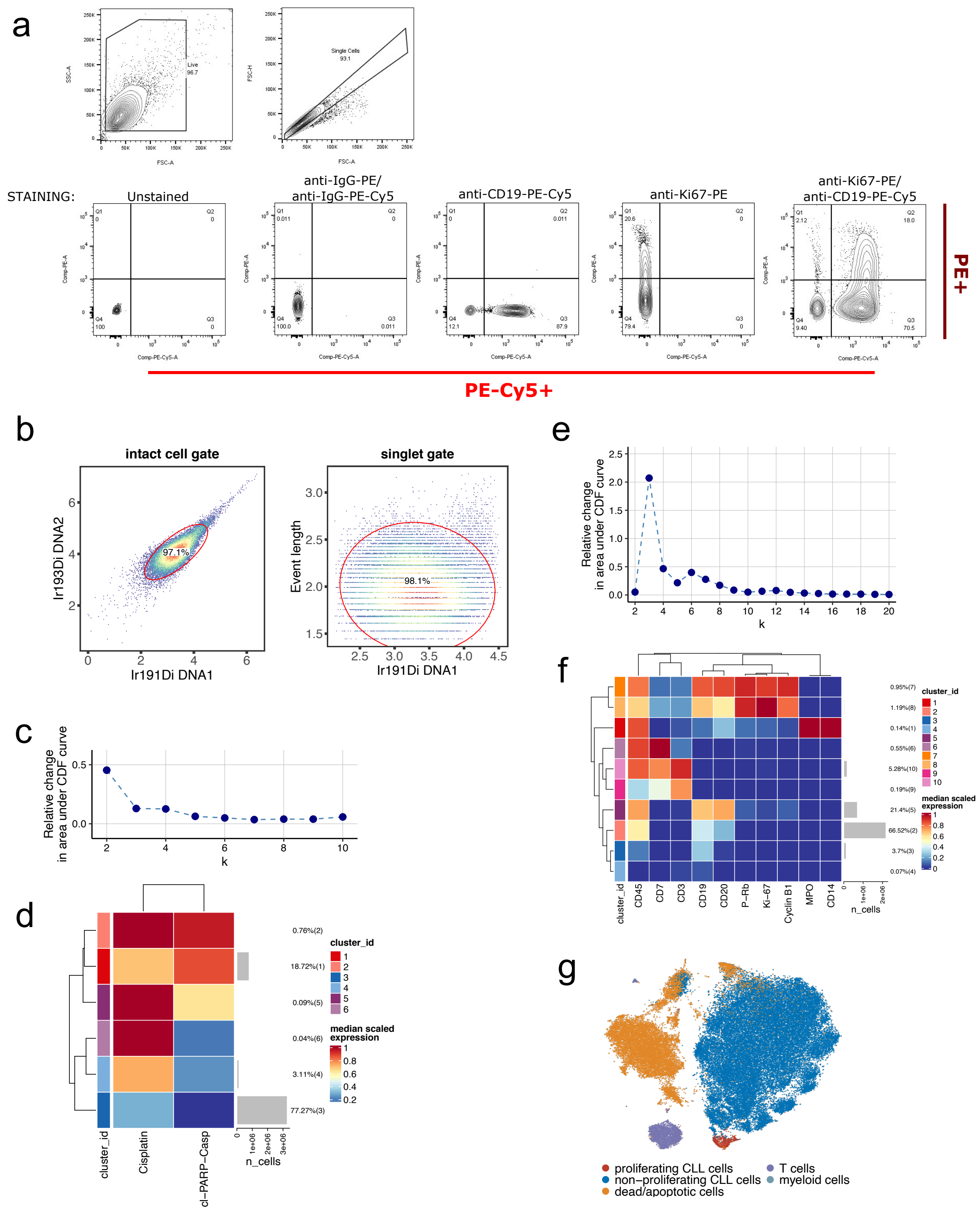


Extended data figure 8. Characterization of CLL-PD by proteomic, ex-vivo drug response and bioenergetic profiling. a, Correlations between CLL-PD to the protein levels of three MYC direct targets that are involved in cell proliferation: *MCM4*, *NME1* and *PAICS*. Per-test *P* values and coefficients are from two-sided Pearson's correlation tests (*n*=46 samples). **b,** *P* values of associations between drug responses and F1 (IGHV), F2 (trisomy12) and F4 (CLL-PD). *P* values are from ANOVA tests including F1, F2 and F4 as covariates. Dashed horizontal line indicates the threshold associated with a false discovery rate (FDR) of 5% (method of Benjamini and Hochberg). **c,** Scatter plots showing the correlations between cell viabilities after drug treatment (averaged over five concentrations tested) and the CLL-PD values. *P* values were from the same ANOVA test as shown in panel b. Only the drugs that showed significant correlations (5% FDR) are shown here. (panel **b** and **c** *n*=190 samples): **d,**

Scatter plots showing the associations of CLL-PD to the three bioenergetic features related to oxidative phosphorylation. Per-test P values and coefficients were from two-sided Pearson's correlation tests (*n*=136 samples). **e,** A heatmap plot showing the z-score of the expression values of proteins that are significantly correlated with CLL-PD (5% FDR, method of Benjamini and Hochberg). Samples (columns) are ordered by their CLL-PD values. The names of mitochondrial proteins are colored in red. **f,** The correlation between the CLL-PD values of 10 samples and their mitochondrial biomass, analyzed by MitoTracker staining. MitoTracker Green (ThermoFisher Scientific, M7514) was used according to the compound's manual. *P* value and coefficient are from two-sided Pearson's correlation tests. **g,** Correlations between CLL-PD and the expressions of two mitochondrial marker proteins, VDAC1 and HSPD1 (HSP60). Per-test *P* values and coefficients in are from two-sided Pearson's correlation tests (*n*=46 samples).



Extended data figure 9. Characterization of CLL-PD at single cell level using CyTOF.
a, The same *t*-SNE layout as shown in Figure 5b, colored by the scaled intensity the other two proliferation markers, P-Rb and Cyclin B1. **b,** A volcano plot showing the differentially expressed markers between CLL-PD high and CLL-PD low samples upon CpG ODN treatment. Text label colors indicate pathway: orange—MYC, purple—mTOR, magenta—BCR, black—other. The y-axis shows the per-test *P* values, which were calculated by differential expression test (based on two-sided moderated t-test) implemented in the diffcyt R package. The dashed horizontal line indicates the threshold associated with a false discovery rate (FDR) of 10% (method of Benjamini and Hochberg) (*n*=8 tumor samples for each of the CLL-PD high and low groups).



Extended data figure 10. **Illustrations of gating and cell type assignment strategies for flow cytometry and CyTOF analyses. a,** Gating strategy used in the assessment of proliferation by flow cytometry. Debris was excluded by gating the largest events based on the side and forward scatter of cells (SSC-A/FSC-A plot). Single cells were selected based on comparison of FSC-H and FSC-A parameters. Ki67+/CD19+ double positive cells were gated among all events based on unstained and staining controls conditions (anti-IgG-PE/anti-IgG-PE-Cy5 isotype controls, anti-CD19-PE-Cy5 and anti-Ki67-PE single staining controls). **b to g,** An illustration of the gating and clustering strategy to annotate cell types in the CyTOF data. **b,** Intact cells and singlets were gated based on the two DNA channels and the event length channel. **c,** Intact cells and singlets were clustered using flowSOM, based on the cisplatin (dead) and cleaved PARP/Caspase3 (cl-PARP-Casp) channels. The number of clusters (k = 6) was chosen based on the elbow point of the relative change in area under CDF curve. **d,** Cells in the cluster that was negative for cisplatin and cl-PARP-Casp (Cluster3) were classified as live cells. Cells in other clusters were classified as dead/apoptotic cells. **e,** Live cells were clustered into 10 clusters using flowSOM based on the intensity of cell lineage and proliferation markers. **f,** Cluster 1, which was positive for CD45, MPO and CD14, was annotated as myeloid cell cluster. Cluster 6, 9 and 10, which were positive for CD45 and CD3 or CD7, were annotated as T cell clusters. Cluster 2, 5, 7 and 8, which were positive for CD45 and CD19, were annotated as CLL clusters. Cluster 3 and 4, which were negative for CD45, may represent non-lymphocytic cells or unhealthy cells and therefore were annotated as dead/apoptotic clusters. Among CLL clusters, Cluster 7 and 8, which are positive for all three proliferation markers, Ki-67, P-Rb and Cyclin B1, were annotated as proliferating CLL clusters, and other CLL clusters were annotated as non-proliferating CLL clusters. **g,** Visualization of cell types on a *t*-SNE map. Due to their low population size (0.14%), myeloid cells are not apparent.